Research on Bearing Intelligent Diagnostic Technology Based on EMD and GA-BP

Zhikai Xing 1, Qiang Wang 1,*, Yongbao Liu 1, Yansong Guo 2, Yan Jia 1, Mo Li 1, and Jun Li 1

1 Department of Power Engineering, Naval University of Engineering, Wuhan 430000, China
2 91998 troops, Lvshun, Dalian116041, China

Abstract. Bearings are the core components of rotating machinery and can have a significant impact on the equipment in the event of a failure. In this paper, an intelligent diagnostic technique based on the combination of EMD and GA-BP algorithm sifts with the rolling bearing fault identification and classification problem. First, the test data is processed by EMD method, the characteristic enhancement and extraction of micro-faults is realized, and the bearings are trouble shoot as training sets and test sets of BPNNs optimized by the built GA. The results show that the accuracy and convergence speed of this method are improved compared with the method of unutilized energy characteristics, and the identification and diagnosis of bearing fault can be effectively carried out.

1 Introduction

Marine gas turbine bearings are often in high temperature, high pressure, strong corrosion, high speed of the harsh working environment, very easy to fail, the use of gas turbine maintenance reliability and economy has a certain impact. [1]

Scholars at home and abroad have done a lot of research on the fault diagnosis of bearings. Zhigang Chen [2] and others combined with the lead convolution network and experience wavelet transformation, improve the ability to extract faults; Zhu Qu [3] and others combined with phased array technology to make up for the lack of vibration monitoring dimension; Xie [4] and others are used to extract deep representation strains from rotating machinery and to identify different faults as training for network models.

In this paper, an EMD-based GA-BPNN fault identification classification method is proposed, which can accurately and quickly classify and identify vibrational signals with noise or interference. The method is easy to identify early weak bearing failure, and early warning and risk-averse, is conducive to early detection of faults, reduce equipment losses.

2 EMD-based GA-BPNN fault identification classification method

The main research ideas and methods in this paper are shown in Figure 1, first the original signal IMF decomposition, calculate the energy characteristics of the decomposition signal, select a higher proportion of energy contribution rate, build the training and test set of the network, build the GA optimization of the BPNN, through the GA iteration to calculate and optimize the weight and threshold in the NN layers.
2.1 Experience modal decomposition principle

The empirical modal decomposition method [5] is a signal processing method proposed by Huang et al. in 1998, the main idea is to break down the original signal into a limited amount of modal function (Intrinsic Mode Function, IMF) and the sum of the residual.

The original signal can be decomposed into the sum of IMF component and a remainder:

\[ x(t) = \sum_{i=1}^{k} c_i + r_{k+1}(t) \]

2.2 GA optimizes BPNN principle

BPNN [6] a multi-layer feed-forward NN consisting of input layer, implied layer and output layer. The GA-BPNN [7] and the global search of the GA and the learning ability of the BPNN, optimizes the weight and threshold of the implied layer and output layer, and accelerates the convergence speed of the network.

Step 1: Calculate the population adaptation value, from which to find out the optimal individual; \( y_i \) represents the ideal output of the i node of GA-BPNN; \( o_i \) represents the identification output of the i node of the network; \( k \) represents the number of network output nodes; A represents the coefficient of the fitness function of an individual; the calculation formula for the fitness value (FIT) of an individual is:

\[ FIT = A \times \left( \sum_{i=1}^{k} |y_i - o_i| \right) \]

Step 2: Select the action; Using to represent the \( \text{FIT}_i \) suitability value of each individual, represents the Nnumber of individual populations, selects according to the individual adaptation ratio, and the probability of each \( f_i \) individual's choice is calculated as:

\[ f_i = \frac{A}{\text{FIT}_i}; \quad p_i = \frac{f_i}{\sum_{j=1}^{N} f_j} \]

Step3: Cross operation; From the method of real crossing, b is a random number in the interval from 0 to 1, and the crossover operation of \( a_x \) on the x chromosome and \( a_y \) on the y chromosome at the z position is as follows:
\[
\begin{aligned}
  a_{xz} &= a_{xz}(1 - b) + a_{yz}b \\
  a_{yz} &= a_{yz}(1 - b) + a_{xz}b
\end{aligned}
\]

Step 4: Determine whether evolution is over, and if not, return to Step 1. The operation parameters are determined by the GA in the initialization, mainly including population size \(M\), genetic algebra \(G\), crossover probability \(p_c\) and mutation probability \(p_m\).

## 3 Experimental data analysis processing

### 3.1 Sorting and grouping data results

The data used in this paper is bearing experimental data from Case Western Reserve University. The test bench includes a 2 hp electric motor, a torque sensor and a control motor. The acceleration sensor used in the experiment is installed vertically on the bearing seat of the bearing to be tested, with a sampling frequency of 12kHz. Select a sensor signal with a fault size of 0.36mm, a sampling frequency of 12kHz, an empty load, and a speed of 1797rpm.

### 3.2 Pre-processing by decomposition signal vibration energy method

The IMF component of various types of failures is obtained through empirical decomposition of each set of experimental data. To better reflect its data characteristics, get the energy of each IMF component and margin of each group:

\[
E_i = \int_{-\infty}^{+\infty} |c_i(t)|^2 \, dt 
\]

Using a group of inner bearing failures as an example, the correlation coefficients of the main components of each IMFs and their energy ratio are calculated as shown in the table. The energy distribution of the other three fault states is concentrated on the 1\(^{\text{ST}}\) IMF component while higher-order IMFs shows some similarity, but their respective energy distributions are not the same. The energy ratio of the IMF component of each set of data as a characteristic input vector provides the basis for the GA-BPNN to classify the fault lines of gas turbine bearings.

**Table 1.** Randomly extract a set of pre-processed results.

<table>
<thead>
<tr>
<th>name</th>
<th>standard deviation</th>
<th>correlation coefficient (%)</th>
<th>energy ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMF1</td>
<td>0.1867542</td>
<td>89.78%</td>
<td>0.8107264</td>
</tr>
<tr>
<td>IMF2</td>
<td>0.0752460</td>
<td>37.88%</td>
<td>0.1316177</td>
</tr>
<tr>
<td>IMF3</td>
<td>0.0326966</td>
<td>10.34%</td>
<td>0.0248509</td>
</tr>
<tr>
<td>IMF4</td>
<td>0.0302347</td>
<td>11.38%</td>
<td>0.0212494</td>
</tr>
<tr>
<td>IMF5</td>
<td>0.0181068</td>
<td>5.35%</td>
<td>0.0076213</td>
</tr>
<tr>
<td>IMF6</td>
<td>0.0107152</td>
<td>2.93%</td>
<td>0.0026697</td>
</tr>
<tr>
<td>IMF7</td>
<td>0.0054010</td>
<td>0.46%</td>
<td>0.0006784</td>
</tr>
<tr>
<td>IMF8</td>
<td>0.0034792</td>
<td>0.04%</td>
<td>0.0002816</td>
</tr>
<tr>
<td>IMF9</td>
<td>0.0022627</td>
<td>0.02%</td>
<td>0.0001190</td>
</tr>
<tr>
<td>IMF10</td>
<td>0.0018410</td>
<td>0.03%</td>
<td>0.0000788</td>
</tr>
<tr>
<td>IMF11</td>
<td>0.0021393</td>
<td>0.03%</td>
<td>0.0001604</td>
</tr>
<tr>
<td>IMF12</td>
<td>0.0001306</td>
<td>0.18%</td>
<td>0.0000004</td>
</tr>
</tbody>
</table>
In the selection of neural network input dimension, if the dimension of the feature vector of the neural network input is too small, it can lead to the entire network model not having characteristic, resulting in a decline in the accuracy of fault diagnosis. If the dimension of the feature vector of the neural network input is too large, it will lead to the entire network model being too complex, and the time of the network model's time reverse solution being too long. In order to reduce the calculation based on ensuring the training accuracy, the first 6-order energy ratio is selected to construct the characteristic vector of failure.

### 3.3 GA-BPNN Solution

According to the classification of faults, each class randomly extracts 180 sets of data, a total of 720 sets of data to form the training set, and the remaining 176 sets of data are numbered, forming the NN's test set. The dimension of the input characteristic vector of the NN is 6, the output vector number is 4, and the implied layer node is determined by the experience formula to be 13. The GA parameters are set: population size is 80, the number of evolutions is 100 and the cross probability is 0.2, the probability of variation is 0.01.

![Figure 2. The optimal individual adaptability value changes.](image)

The optimal threshold and optimal weight calculated by the GA algorithm are brought into the BPNN, the input characteristic vector of the test set is then input into the GA-BP network, to compare the actual recognition results with the theoretical output results.

In the test set of the two networks, 40 groups were randomly selected, compared with their ideal identification classification and actual identification classification, vertical axis represents the category of fault, 1 indicates the inner ring fault, 2 indicates the outer ring fault, 3 indicates the normal state, 4 indicates the rolling body fault.

![Figure 3. The recognition of a BPNN with an energy signal as a characteristic input.](image)
Calculate the number of predicted outputs for all test sets and record the calculation time. The algorithmic accuracy and the calculation time pair are shown in the table below.

<table>
<thead>
<tr>
<th>Input characteristic vectors of NNs</th>
<th>Vibration value of the original signal</th>
<th>Energy values for each component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation accuracy of training samples</td>
<td>0.6817</td>
<td>0.0047</td>
</tr>
<tr>
<td>Time required to solve</td>
<td>0.6582</td>
<td>0.2888</td>
</tr>
<tr>
<td>Accuracy within the confidence interval of 0.95</td>
<td>91.48%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Accuracy within the confidence interval of 0.98</td>
<td>81.82%</td>
<td>97.73%</td>
</tr>
</tbody>
</table>

Using the data in the table, the bearing fault identification network based on the EMD and GA-BP algorithms has an accuracy of 97.73%, much larger than the accuracy of the original signal. It can be noted that when the original data is processed in combination with EMD, the algorithm accuracy is greatly reduced and the calculation time is reduced by half.

### 4 Conclusions of the paper

To avoid recognition errors caused by similar vibration amplitudes, EMD is introduced into early data pre-processing. On this basis, the NN is constructed in combination with the GA-BP algorithm. The rationality of the method is proved by the data analysis of bearing experiments. The main conclusions are as follows:

1. Compared with the original vibration signal, the application of EMD has higher recognition ability to small outliers and similar vibration values in the original signal. This method can solve the problem of vibration signal instability, overcome the energy leakage problem of wavelet analysis, have a better recognition effect on similar vibration values.

2. Compared with other algorithms, the bearing fault identification method based on EMD and GA-BP greatly improves its convergence speed and convergence accuracy, improves the weight and threshold setting of NN, appropriately enlarges the weak bearing fault characteristics, and has a more far-reaching use for classification recognition.

### References